Kap. 8: Utvalsfordelingar og databeskrivelse

Utvalsfordelingar

- Utvalsfordeling for gjennomsnitt (med kjent varians) ($\bar{X}$)
- Sentralgrenseteoremet (SGT)
- Utvalsfordeling for varians (normalfordeling)
- Utvalfordeling for gjennomsnitt (normalfordeling og ukjent varians)

Databeskrivelse:

- Mest i øving
- Jf første veka
- qq-plott (er dataene frå ei normalfordeling?)

Observatorar

Frå eit datasett finn vi (observerer vi) typisk:

Empirisk gjennomsnitt: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Empirisk varians: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Empirisk median: $\tilde{x} = x_{(\frac{n+1}{2})}$ (dersom n odde)

Observatorar

Frå eit datasett finn vi (observerer vi) typisk:

Empirisk gjennomsnitt: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Empirisk varians: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Empirisk median: $\tilde{x} = x_{(\frac{n+1}{2})}$ (dersom n odde)

Om vi ser på desse teoretisk (desse er observatorar):

Gjennomsnitt: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Utvalsvarians: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Median: $\tilde{X} = X_{(\frac{n+1}{2})}$ (dersom n odde)

Observatorar

Frå eit datasett finn vi (observerer vi) typisk:

Empirisk gjennomsnitt: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Empirisk varians: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Empirisk median: $\tilde{x} = x_{(\frac{n+1}{2})}$ (dersom n odde)

Om vi ser på desse teoretisk (desse er observatorar):

Gjennomsnitt: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Utvalsvarians: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Median: $\tilde{X} = X_{(\frac{n+1}{2})}$ (dersom n odde)

Def. 8.4 Observator

Ein funksjon av stokastiske variable som representerer eit tilfeldig utval blir kalla ein observator (engelsk: statistic) $\Rightarrow$ har sannsynsfordeling.

Ein observator er ein stokastisk variabel, då det er ein funksjon av stokastiske variable.

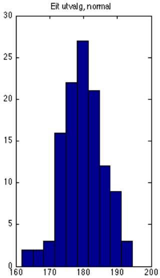
Utvalgsfordelinga til $\bar{X}$

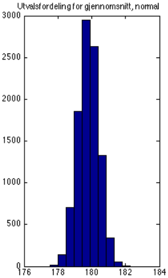
Algoritme

For $m = 1 : M$

- Trekk n=117 datapunkt frå $N(179.8, 6.5^2)$ $m = 1, \ldots, M$ gongar $\Rightarrow x_{m1}, x_{m2}, \ldots, x_{mn}$.
- Finn gjennomsnittet $\bar{x}_m = 1/n \sum_{i=1}^{n} x_{mi}$

Plott histogram for $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M$





Dersom X_i er normalfordelt, og kjent varians

$X_i \sim N(\mu, \sigma^2)$, for $i = 1, 2, \ldots n$. Dvs eit utvalg på n.
Då er

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Sentralgrenseteoremet, teorem 8.2

La X_i, $i = 1, 2, \ldots, n$ vere uavhengige identisk fordelte (u.i.f.) stokastiske vaiable med $E(X_i) = \mu$ og $Var(X_i) = \sigma^2 < \infty$ (endeleg varians). La $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ og $Z = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}$. Då

$$Z \to N(0, 1)$$

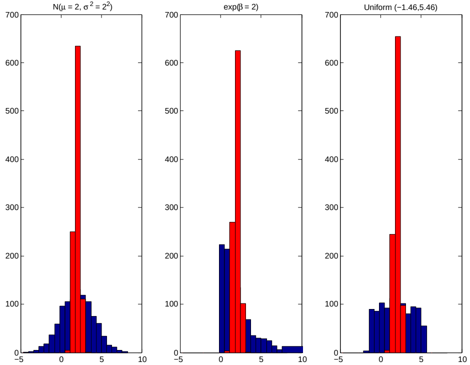
når $n \to \infty$.
Dette tilsvarer

$$\bar{X} \to N(\mu, \frac{\sigma^2}{n})$$

når $n \to \infty$.

PS: Gjeld uansett fordeling for X_i.
PSS: $Z = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}$ er og ein observator

Histogram populasjon og gj.snitt av 30



I dag

Skal sjå på:

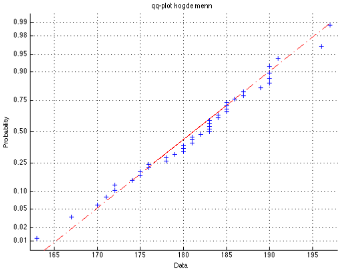
- $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
- $T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$

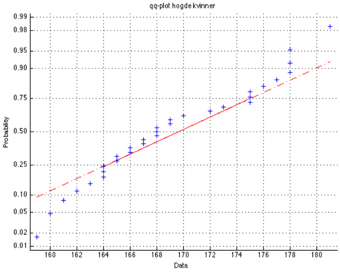
Gjeld kun for $X_1, \dots X_n \sim N(\mu, \sigma^2)$

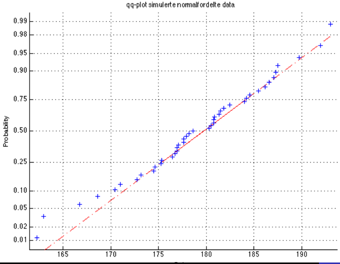
- Korleis finn vi ut om dataene våre er normalfordelte?

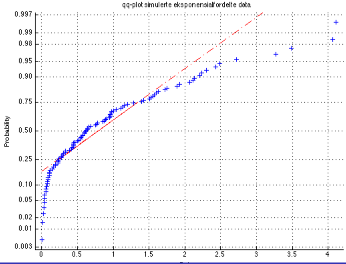
Kan bruke kvantil-kvantil-plott (qq-plot, i Matlab *normplot*)
qq-plot: Dersom normalfordelt, ca på rett linje. Sjå og `http://www.math.ntnu.no/emner/TMA4240/2012h/Kompendium/notater/qqplott.pdf`, notat om qq-plot. Eksempla er det viktige!

qq-plot

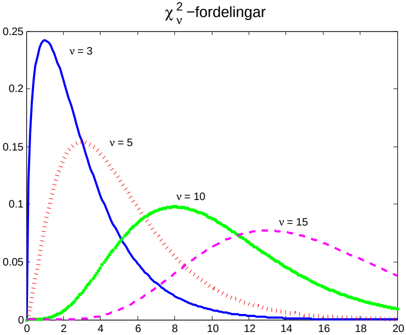

qq-plot hogde menn
Probability
Data


qq-plot hogde kvinner
Probability
Data

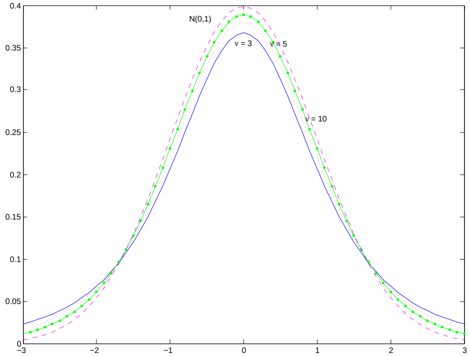

qq-plot simulerte normalfordelte data
Probability


qq-plot simulerte eksponensialfordelte data
Probability

χ^2-fordelingar



Student t-forelingar



Oppsummering

Utvalsfordelingar

- Utvalsfordeling for gjennomsnitt (med kjent varians)
- Sentralgrenseteoremet (SGT) uansett fordeling
- Utvalsfordeling for varians normalfordeling, χ^2_{n-1}
- Utvalfordeling for gjennomsnitt normalfordeling og ukjent varians, T_{n-1}

Databeskrivelse: boksplott og qq-plot